DOCUMENT RESUME

ED 391 814 TM 024 153

AUTHOR Stocking, Martha L.
TITLE Some Considerations in Maintaining Adaptive Test Item Pools.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.
REPORT NO ETS-RR-88-33-ONR
PUB DATE May 88
CONTRACT N00014-85-K-0683
NOTE 60p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Adaptive Testing; Comparative Analysis; *Computer Assisted Testing; Identification; *Item Banks; Psychometrics; Scores; *Selection; Simulation; Test Construction; Test Format; *Test Items; Test Results
IDENTIFIERS *Parallel Test Forms; Test Security

ABSTRACT

The construction of parallel editions of conventional tests for purposes of test security while maintaining score comparability has always been a recognized and difficult problem in psychometrics and test construction. The introduction of new modes of test construction, e.g., adaptive testing, changes the nature of the problem, but does not make it disappear. Items in adaptive testing pools may become overused and require replacement. However, in order to insure score comparability, important characteristics of the pool must remain constant. Three methods of selecting candidate new items and three methods of identifying items for replacement are developed and compared with each other and with a previous method through a simulation study. Results indicated that using conventional item statistics to screen items before deciding to seed them was important and effective in terms of maintaining the information structure of the adaptive test item pool. The online calibration of larger sets of seeded items from which to select replacements can substantially improve the ease with which the information structure of the pool can be maintained. (Contains 1 table, 11 figures, and 4 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 391 814

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SOME CONSIDERATIONS IN MAINTAINING ADAPTIVE TEST ITEM POOLS

Martha L. Stocking

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-85-K-0683

Contract Authority Identification No.
NR No. 150-539

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, New Jersey

May 1988

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

Approved for public release; distribution unlimited.

TM024153

BEST COPY AVAILABLE

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a REPORT SECURITY CLASSIFICATION: Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S): RR-88-33-ONR	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION: Educational Testing Service; 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION: Cognitive Science Program, Office of Naval Research (Code 1142CS), 800 North Quincy Street
6c ADDRESS (City, State, and ZIP Code): Princeton, NJ 08541	7b ADDRESS (City, State, and ZIP Code): Arlington, VA 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION; 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-85-K-0683
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO 61153N; PROJECT NO RR04204; TASK NO RR04204-01; WORK UNIT ACCESSION NO NR150-539

11 TITLE (Include Security Classification): Some Considerations in Maintaining Adaptive Test Item Pools (Unclassified)

12 PERSONAL AUTHOR(S): Martha L. Stocking

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Technical	FROM _____ TO _____	May 1988	47

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Adaptive testing; Item pool refreshment; On-line calibration; Parallel adaptive tests
05	10		

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The construction of parallel editions of conventional tests for purposes of test security while maintaining score comparability has always been a recognized and difficult problem in psychometrics and test construction. The introduction of new modes of testing, e.g., adaptive testing, changes the nature of the problem but does not make it disappear. Items in adaptive test item pools may become overused and require replacement. However, in order to insure score comparability, important characteristics of the pool must remain constant. Three methods of selecting candidate new items and three methods of identifying items for replacement are developed and compared with each other and with a previous method through a simulation study.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL: Dr. James Lester	22b TELEPHONE (Include Area Code): 202-696-4503; 22c OFFICE SYMBOL: ONR 1142CS

DD Form 1473, JUN 86 — Previous editions are obsolete — S/N 0102-LF-014-6603

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

Some Considerations in Maintaining Adaptive Test Item Pools

Martha L. Stocking

Educational Testing Service

May 1988

Copyright © 1988. Educational Testing Service. All rights reserved.

Abstract

The construction of parallel editions of conventional tests for purposes of test security while maintaining score comparability has always been a recognized and difficult problem in psychometrics and test construction. The introduction of new modes of testing, e.g., adaptive testing, changes the nature of the problem but does not make it disappear. Items in adaptive test item pools may become overused and require replacement. However, in order to insure score comparability, important characteristics of the pool must remain constant. Three methods of selecting candidate new items and three methods of identifying items for replacement are developed and compared with each other and with a previous method through a simulation study.

Keywords: Adaptive testing
On-line calibration
Item pool refreshment
Parallel adaptive tests

Some Considerations in Maintaining Adaptive Test Item Pools

Introduction

Test development specialists and psychometricians have long struggled with the problems associated with the construction of parallel editions of a single conventional test. The decision to issue a new test edition is usually based on the desire to preserve test security by preventing overexposure of test editions. Typically, all items in a conventional test are replaced by a new set of items that conform to the same content and statistical specifications as the original test edition. To compensate for any remaining differences between the new and original test editions, statistical procedures are usually employed to insure that scores resulting from the administration of either test edition have the same interpretation.

New advances in psychometrics and computer technology encourage individualized (adaptive) testing on a microcomputer, where each examinee is administered a small set of items drawn from a larger item pool. Using a possibly very complex set of decision rules, examinees may receive completely different sets of items. Two issues immediately arise in this context. First, in order to make examinee scores comparable on different sets of items, measures must be taken to control the content and

statistical properties of the item sets appropriately. Second, when faced with decisions to replace overexposed items in the item pool from which the individualized tests are drawn, care must be taken to insure that the characteristics of the item pool remain as nearly constant as possible, so that the accuracy of estimated adaptive test scores remains the same across various editions of the item pool. Issues surrounding this latter topic are addressed in this paper.

The next section describes an idealized setting for adaptive testing as a context for some practical constraints. A convenient method of analyzing and comparing certain features of item pools is detailed in the following section. Remaining sections of this paper will describe a particular practical problem in maintaining adaptive test item pools, and some potential solutions to this problem. An investigation of the efficacy of these solutions when applied to simulated data is described, and the results discussed.

An Idealized Setting and Some Practical Constraints

The major psychometric appeal of adaptive testing is the promise of equally precise measurement of all examinees, regardless of their ability levels. Aside from the details of a particular adaptive testing algorithm, the promise of equal measurement precision rests on certain strong assumptions about

the item pool. The first assumption made is that it is possible to obtain sufficient numbers of items appropriate for all ability levels. Secondly, it is assumed that the 'appropriateness' of an item is related to the precision with which a particular item will measure an examinee with a particular level of ability. The third assumption made is that the set of items appropriate for a particular level of ability represents a certain average level of precision, and that this precision remains constant across examinee ability levels. In the circumstances considered in this paper in which items in the pool must be replaced from time to time, it is further assumed that the replacement items are psychometrically equivalent to the items being discarded.

Thus, in an idealized setting in which the goal of testing is to measure all abilities with equal precision, the ideal item pool consists of sufficient numbers of items whose measure of precision follows a rectangular distribution across the entire ability range to be measured. Further, in this setting, the psychometric properties of this ideal item pool are not affected by the process of discarding some items and replacing them with others. Given sufficiently expert item writers, with sufficient time and money to complete many cycles of item writing and pretesting, it is possible that this ideal situation could be realized.

However, in practice, many compromises are made: 1) The abilities of interest are restricted to some finite range. This automatically decreases the necessary item production effort by denoting ability levels outside the specified range as unimportant.

2) The size of the item pool is limited. The limit is determined not only by the numbers of items required for adaptive tests of various lengths, but also by the computer resources required for item storage and display.

3) In the production of items for the pool, only a finite number of cycles of item writing and pretesting are conducted. Thus the item pool will consist of the best items that could be produced for a certain fixed cost. It is unlikely that such a pool can contain sufficient numbers of appropriate items, even for the abilities within the restricted range of interest. Thus a further compromise is required -- to measure some ability levels with more precision than other ability levels.

4) If the adaptive test is administered to a group of examinees whose distribution of ability is bell shaped, the items most vulnerable to overexposure, for commonly used item selection algorithms, are those that are most appropriate for the average examinee. In the production of replacement items, only a finite

number of cycles of item writing and pretesting are conducted, as before. Even with the most sophisticated item writers, it is unlikely that this production effort can be sufficiently narrowly focused to result in an adequate number of items that are psychometrically equivalent to those items most appropriate for the average examinee. Thus another compromise -- the psychometric properties of the item pool may change over cycles of item pool refreshment.

5) Some items in the item pool may be appropriate for such extreme ability levels that they are infrequently, and sometimes never, administered when the adaptive test is given to finite samples of examinees. This naturally leads to the consideration of removing these items, to gain more room in an item pool of fixed size for items that are appropriate for more typical examinees. In the real-world situation, where items are appropriate at more than a single level of ability, this can be a mechanism for increasing precision at typical levels of ability at the sacrifice of precision at more extreme levels of ability. This results in yet another compromise -- the 'effective' range of the abilities of interest is shrunk.

The issues addressed in this paper arise in the context of the constraints and compromises imposed by the process of moving

adaptive testing out of the theoretical realm and into the practical realm.

A Convenient Method of Analyzing Certain Item Pool Features

The adaptive test algorithm used in this paper, as well as most adaptive testing algorithms in current use, rest on modern model-based psychometrics such as Item Response Theory (IRT). In IRT, one way of characterizing the precision with which an item measures an ability is by the item information function (Lord, 1980, equation 5-9). The information structure for a collection of items can be characterized by the test information function (Lord, 1980, equation 5-6), which is formed by taking the simple sum, at different abilties, of the values of the item information functions. This test information function is the maximum amount of information that can be obtained from the item set if it were administered as a conventional test.

It bears emphasizing to note that the test information for an adaptive test item pool <u>is not</u> the information function for an adaptive test using this item pool. The adaptive test information function depends upon the items actually taken by examinees. This is determined not only by the information structure of the item pool, but also by the details of the algorithm such as those that specify the selection of the first and subsequent items for

administration, randomization of item selection to increase item security, the rule used to stop item administration, and the method of scoring the adaptive test. The adaptive test information function for algorithms of the type used here can only be conveniently estimated from numerical approximations using Monte Carlo results (see, for example, Lord, 1980, section 10.6).

In this discussion, the estimated test information function is viewed as a convenient mechanism for discovering changes in the information structure of the item pool upon which the adaptive testing algorithm will operate. This estimated test information function is obtained by substituting estimated, rather than true, parameters into Lord's equation, and is the only test information that is computable in practical applications where true parameters are unknown. In the context of the idealized setting discussed previously, the optimum item pool, in terms of an information measure, would have constant estimated test information across all ability levels, and would not change as items are discarded and replaced.

A Practical Problem in Item Pool Maintenance

A number of agencies of the Department of Defense recently funded a three-year project to develop and evaluate different methods of on-line calibration for the computerized adaptive Armed

Services Vocational Aptitude Battery (CAT-ASVAB) (Bock, Davis, Holland, Levine, Samejima, & Stocking, 1988). On-line calibration methods are procedures to obtain parameter estimates for new items that are candidates for inclusion in subsequent item pools from data collected during an examinee's testing session (on-line). For this particular project the final parameter estimates were constrained to be based on the 3-parameter logistic model of item response functions (Lord, 1980, equation 2-1). As part of this project, a method of on-line calibration based on the estimation procedures in the LOGIST computer program (Wingersky, 1983) was explored by the author; Bock, Levine, and Samejima developed other methods.

In on-line calibration, each examinee is administered (seeded) a small number of items that are candidates for inclusion in the next version of the item pool. In the LOGIST-based method, examinees are also administered a small number of 'anchor' items. These anchor items are not part of the adaptive test item pool, although they have well-determined parameter estimates that are on the same metric as those of the item pool. The responses to neither the seeded items nor the anchor items are used in the operation of the adaptive test algorithm. In the LOGIST-based method of on-line calibration, the responses to items administered

in the adaptive test are used to compute a maximum likelihood estimate of examinee ability. The item responses to the seeded items and the anchor items are used, along with these ability estimates, to obtain parameter estimates for the seeded items and to reestimate the parameters for the anchor items. The two sets of parameter estimates for the anchor items, the original set on the scale of the item pool and those resulting from the on-line response data collection, are used to develop a scaling transformation that places the parameter estimates for the seeded items onto the metric of the adaptive test item pool.

The final phase of the On-line Calibration project consisted of a sequence of four simulations of adaptive testing and item pool refreshment for each method of on-line calibration. The generating (or true) item response functions used in the simulations were nonparametric (and frequently nonmonotonic) functions developed by Levine (Bock et al., 1988). All simulated examinees (simulees) were drawn from a bell-shaped distribution of true ability also generated by Levine (Bock et al., 1988). Davis (Bock et al., 1988) selected seeded items, conducted all simulations of adaptive testing and the collection of data on the seeded items, and identified items already in the pool to be replaced. Individual experimenters were responsible for the on-

line calibration of seeded items and the selection of a subset of these to replace those items to be discarded from the pool.

Starting with an initial item pool (called the Round 0 pool), adaptive testing was simulated using this pool; responses to seeded items were collected simultaneously. Items were then identified for elimination from the pool, and, for the LOGIST-based method, replacement items were selected from the seeded new items to maintain an item pool of constant size with an information function similar to that of the Round 0 pool. This was considered to be the first 'Round' of adaptive testing and item pool refreshment. The second Round proceeded using the refreshed pool created during the first Round; the third Round used the refreshed pool from the second Round; and the fourth and final Round used the refreshed pool from the third Round.

During the progress of these simulations, it became apparent that the rule employed for the selection of candidate new items for seeding and the rule for the elimination of old items from the pool had important impacts on the information structure of subsequent item pools. The original item pool consisted of 100 items that were selected on the basis of estimated information from a collection of 258 5-choice items. At each Round (of four) of adaptive testing and item pool refreshment, a set of 50 items

to seed was obtained by random selection from the collection of 258 items. Also at each Round, the 25 items already in the pool that received the highest number of administrations in the adaptive test simulations, accumulated across the current and all previous Rounds, were designated as items that must be replaced by selecting 25 (half) of the seeded items.

Eliminating the 25 items most frequently used in adaptive test simulations, where simulees were drawn from a typical distribution of true ability, resulted in the elimination of 25 middle difficulty items with good discriminations and low guessing parameters in Round 1. The attempt to replace the eliminated items by selecting half of the 50 seeded items resulted in an initial large decrease in estimated test information for the item pool at middle ability levels on the first Round, and small fluctuations around this initial decrease in subsequent Rounds. Figure 1 shows the estimated test information functions for the item pool at each Round for the LOGIST-based method of on-line calibration.

Insert Figure 1 about here

By changing the rules used to select items for seeding and for elimination from the pool, it should be possible to produce less dramatic changes in the information structure of the item pool. This study tries out three selection rules and three elimination rules.

Selection Rules

During the previous simulations, seeded items were randomly selected from the collection of 258 items. In every Round, the 25-item set selected as replacement items was nearly as good, in terms of estimated test information for middle ability levels, as the complete set of 50 candidate items. Figure 2 shows the estimated test information functions for the set of 50 seeded items and the 25 replacement items selected from them for the refreshment of the Round 0 pool. These results are typical of other Rounds. For improvements in the process for middle ability levels, then, we need to improve the quality of the items selected for seeding.

Insert Figure 2 about here

In practice, items should not be considered as candidates for an adaptive test item pool until some rough idea has been obtained

as to their quality. A reasonable approach would be to gather some conventional statistics on such items for screening purposes. The three rules proposed here utilize the conventional proportions-correct and r-biserials.

Selection Rule 1

Selection Rule 1 will consider only those of the 258 items that have conventional proportions-correct between .2 and .9, and r-biserials of at least .2. Of those items meeting these criteria, a random sample of 50 will be selected as the set of items to be seeded. This rule is only a slight modification of the previous rule.

Selection Rule 2

This Selection Rule will use the same relatively unrestrictive screening of the collection of 258 items, but will randomly select 100 items as the set of items to be seeded. This is a greater departure from the previous rule in that twice as many items are now available for possible selection into the adaptive test pool.

Selection Rule 3

This Selection Rule will use a more restrictive screening. We know that it is the middle difficulty items that will be most used when the adaptive test is administered to a typical group.

It seems reasonable to capitalize on this knowledge. This Selection Rule, like the others, will eliminate those items of the 258 with r-biserials less than .2. Then, 100 items will be selected for seeding whose proportions-correct are between .4 and .8, indicating that these items are about middle difficulty for 5-choice items.

Elimination Rules

At the end of the previous four Rounds of simulations, about 30% of the final item pool consisted of items that were retained from the initial item pool. All of these retained items had difficulties greater than 1.0 in absolute value. Although these items had been available for administration to 60,000 simulees by the end of Round 4, they had not accumulated sufficient responses to be among the 25 most used items at any Round. A different but overlapping 30% of the final item pool consisted of items with fewer than 1000 (and sometimes no) cumulative responses. Most of these items had estimated difficulties greater than 1.5 in absolute value. To retain so many little used items in the face of the change in information structure of the pool for average examinees may be inefficient for adaptive testing with a typical group of simulees. It may be more efficient to shrink the

effective ability range of interest. Two of the three elimination rules proposed here capitalize on this idea.

Elimination Rule 1

This Elimination Rule is identical to that used in the previous study. The 25 items receiving the highest number of adaptive administrations will be eliminated from the pool and replacements selected for them.

Elimination Rule 2

The 25 items most used in the adaptive test simulations will be eliminated, as in Elimination Rule 1. In addition, the 25 items least used in the adaptive test simulations will also be eliminated. A set of 50 replacement items will be selected.

Elimination Rule 3

As before, the 25 most used items will be eliminated. In addition, the 5 least used items will be eliminated and a set of 30 replacement items will be selected.

The Current Study

The Data

For purposes of this study, it was decided to focus on the item pools from Round 0 and Round 1 of the previous simulations. The change in information structure is largest between these pools, which were used for the first and second adaptive test

simulations, respectively. While the method used to build the initial Round 0 item pool produces an overly optimistic estimated test information function for that pool, the change in the characteristics of the pool from Round 0 to Round 1 is real. Figure 3 shows the drop in the true information function for the Round 1 pool when compared to that of the Round 0 pool.

Insert Figure 3 about here

Implementation of Selection Rules

Davis provided the data for the computation of conventional proportions-correct and r-biserials by simulating the administration of all 258 items to a random sample of 500 simulees. These simulees were drawn from the same distribution of true ability used in the previous simulations. Figure 4 shows a scatterplot of the r-biserials against the proportions-correct for all 258 items. Approximately half of the 258 items are easy items with proportions-correct above .9.

Insert Figure 4 about here

There were 118 items that met the criteria for inclusion for Selection Rules 1 or 2, i. e., r-biserials of at least .2 and proportions-correct between .2 and .9. From this set of items, 100 were chosen at random to form the set of Selection Rule 2 seeded items. Of these 100, a randomly chosen subset of 50 were selected to be the seeded items for Selection Rule 1. Summary statistics for both of these item sets are shown in Table 1. For both sets of items, the correlation between proportions-correct and r-biserials is moderately high. This suggests that the more difficult items are also more informative.

Only 51 items met the criteria for inclusion for Selection Rule 3. To provide the necessary 100 items, 49 items were sampled randomly with replacement from the 51. Summary statistics for Selection Rule 3 items are also shown in Table 1. The correlation between proportions-correct and r-biserials is reduced as are standard deviations when compared to the other item sets because the range of proportions-correct is restricted.

Insert Table 1 about here

Adaptive Test Simulations

Davis simulated the administration of an adaptive test to a sample of 30,000 simulees drawn from the distribution of ability used in the previous study. In addition to the adaptive test, each simulee responded to a random set of five anchor items (out of 25) as required by the LOGIST-based method of on-line calibration. Each of the first 15,000 simulees was seeded a random set of five of the 50 Selection Rule 1 items. All 30,000 simulees were seeded random sets of five of the 100 Selection Rule 2 items and also random sets of five of the 100 Selection Rule 3 items. Thus each anchor item received about 6000 responses, and each of the items in the sets of seeded items received about 1500 responses.

On-line Calibrations

Three separate on-line calibrations were preformed using the LOGIST-based anchor item approach developed for the previous study, one for each set of seeded items associated with a particular Selection Rule. The first LOGIST calibration used the 15000 simulees who responded to Selection Rule 1 items as well as the anchor items. The second LOGIST calibration used the 30,000 simulees responding to Selection Rule 2 items as well as the anchor items. The final LOGIST calibration used the same 30,000

simulees, but only their responses to the anchor items and the Selection Rule 3 item set. Characteristic curve scale transformations (Stocking and Lord, 1983) using the new item parameter estimates for the anchor items were then used to place the results of each calibration, independently, onto the scale of the Round 0 item pool.

The Choosing of Replacement Items

The Elimination Rules studied mandate the discarding of 25, 50, or 30 items from the pool. The Selection Rules prescribe the choice of sufficient items to maintain pool size from a set of 50 or one of two different sets of 100 candidate new items. Regardless of the number of items to be discarded or the set from which replacements were to be selected, the same algorithm was used to choose the appropriate number of replacement items from the set of seeded items.

A 'target' information function was defined as the estimated test information function of the items discarded using Elimination Rule 1, that is, the 25 items most frequently used in the adaptive test simulation. The use of this target across Selection and Elimination Rules insures that the space obtained in the pool by discarding any little-used items will be utilized to select replacement items for the over-used items only.

Two methods of choosing items to match the target information function were employed. The first method chose items with the greatest area under their estimated item information functions within ability levels that appeared important based on the target information function. The second method chose items on the basis of the area under the estimated item information functions and then attempted to improve on this by discarding some items and selecting others that minimized the maximum difference between the target and the draft estimated test information functions.

Neither of these methods of choosing replacement items worked automatically without intervention. The replacement items were ultimately chosen on the basis of a subjective criterion: item sets with estimated information functions closer to the target over middle ranges of ability were preferable to item sets with estimated information functions more distant from the target in the middle but closer at the extremes. Both of the methods required tinkering with the ability limits within which a match to the target was desired.

Results

The sets of seeded items resulting from each Selection Rule were used with each Elimination Rule to develop a new 100 item pool. That is, the 50-item set of seeded items resulting from

Selection Rule 1 was used as a source of 25, 50 and 30 replacement items for Elimination Rule 1, 2 and 3 respectively. The same pattern was repeated for the 100 item sets resulting from Selection Rules 2 and 3. The effects of the different Selection and Elimination Rules were compared to each other through the use of the estimated test information function for the resulting 100 item pool. These results were also compared to the original Round 0 pool estimated test information function, as well as to the previous Round 1 pool estimated test information function.

Figure 5 shows the estimated test information functions for the sets of seeded items resulting from the three Selection Rules. These can be interpreted as showing what is available to work with, in terms of estimated information, when selecting the appropriate number of replacement items for each Elimination Rule. Also on the same plot is the target test information function for the 25 most used items in the Round 0 item pool. As expected, the estimated information function for Rule 3 is highest and narrowest; the conventional proportions-correct for the items selected covered a fairly narrow range. Also as expected, the shapes of the estimated information functions for Selection Rules 1 and 2 are similar, with the Rule 2 estimated information function about twice as high as that for Rule 1 in the middle

ranges of abilities. This seems reasonable since the same moderate screening was applied under both rules, and there are twice as many Rule 2 items as Rule 1 items.

Insert Figure 5 about here

Also shown on the same figure is the estimated test information function resulting from the random selection rule used in the previous simulations. While the Rule 1 set has the same number of items, it is clearly a more informative set of items than that chosen by the previous random selection rule.

Selection Rules

Figure 6 displays the results for Selection Rule 1 using each Elimination Rule, in terms of estimated information (top) and relative efficiencies (bottom) of the resultant 100-item pools. The estimated information for the original Round 0 pool and the Round 1 pool produced using the random selection rule and elimination rule of the previous study are displayed on the graph for comparison. It seems clear that the moderate screening is effective. Replacing the 25 most used items (Elimination Rule 1) with 25 items that have been subjected to a moderate screening yields a higher estimated information function for middle ability

levels than if the 25 items have not been screened. Replacing the 25 most used and 25 least used items in the pool (Elimination Rule 2) with all 50 of the moderately screened items is less satisfactory. The estimated information is only slighted higher than when replacing 25 items in the middle, but too high at higher abilities and too low at lower abilities. Replacing 30 rather than 25 items (Elimination Rule 3) is only a very slight improvement over replacing 25 items.

Insert Figure 6 about here

The same conclusions may be drawn from the relative efficiency graph. The efficiency of each pool constructed by the present rules and the previous rule is computed relative to the Round 0 pool.

Comparable plots of estimated information and relative efficiencies are displayed for Selection Rule 2 (Figure 7) and Selection Rule 3 (Figure 8). Selection Rule 2, which provides 100 moderately screened seeded items, does substantially better than random selection. This is true even when only the 25 most used items in the Round 0 pool are identified for replacement (Elimination Rule 1). When 30 items are to be replaced

(Elimination Rule 3), Selection Rule 2 produces a new pool that has nearly the same information structure as the original Round 0 pool. When 50 items are to be replaced -- the 25 most used and the 25 least used (Elimination Rule 2) -- Selection Rule 2 produces a pool that has higher test information for middle ability levels, and lower test information for extreme ability levels, when compared to the Round 0 pool.

Insert Figures 7 and 8 about here

Selection Rule 3, by providing 100 seeded items that have been subjected to a more restrictive screening, nearly matches the information structure of Round 0 pool when either 25 or 30 items are replaced (Eliminations Rules 1 and 3). When 50 items are replaced (Elimination Rule 2), the resultant new pool's information structure is changed to be sharply higher at middle ability levels and lower at extreme ability levels.

Elimination Rules

The three Figures just examined show, for each Selection Rule, the consequences of each Elimination Rule. It is also informative to look at the same data along the other facet, that is, for each Elimination Rule, the consequences of each Selection

Rule. Figures 9 through 11 show the new pools produced by each Selection Rule for Elimination Rules 1, 2 and 3 respectively.

Insert Figures 9, 10 and 11 about here

When only the 25 most used items are eliminated (Elimination Rule 1, Figure 9), seeded items from Selection Rule 3 provide the item pool most similar to the Round 0 pool. The results of all selection rules are, in fact, strictly ordered for middle ability levels in terms of information structure. The most different new pool is produced when the set of seeded items has been randomly selected. The most similar new pool is produced when the set of candidate items is larger, and has been subjected to fairly strict screening. This pool is nearly as good as the Round 0 pool in terms of estimated information for middle ability levels.

When 50 items are to be eliminated (Elimination Rule 2), Figure 10 shows that selecting replacements from the larger item sets (Selection Rules 2 and 3) produces new pools that are more informative than the Round 0 pool for middle ability levels and less informative for more extreme ability levels. This may be undesirable because the information structure of the resultant pools is changed for almost all levels of ability. Selecting as

replacements all 50 seeded items provided by Selection Rule 1 does not yield as much information as the Round 0 pool for middle or low ability levels, but yields more information for higher ability levels.

A more moderate approach is to eliminate the 25 most used and the 5 least used items in the pool (Elimination Rule 3, Figure 11). In terms of reproducing the estimated information structure of the Round 0 pool, selecting 30 items from 100 items that have been moderately screened (Selection Rule 2) or more strictly screened (Selection Rule 3) produce very similar results. Both of these approaches replicate the Round 0 estimated information structure well. Selecting 30 replacement items from only 50 moderately screened seeded items provided by Selection Rule 1 provides more information than the random selection approach, but still does not approximate the estimated information structure of the Round 0 pool very well.

Discussion

The context of this study has been adaptive tests administered to examinees whose distribution of ability is bell-shaped. While this is probably the most common context in which adaptive testing is implemented, it should be noted that the details of the Selection and Elimination Rules studied here might

be inappropriate if the distribution of examinee ability had a very different shape. Consider, for example, the situation in which the distribution of ability is U-shaped rather than bell-shaped. Then the screening on proportions-correct for the Selection Rules considered here eliminates exactly those items that are most likely to be useful.

The criterion used to evaluate the operation of Selection and Elimination Rules was the information structure of a particular item pool. Although this item pool was built by the commonly accepted methods in adaptive testing, this pool would have been different if different items had been available for its construction. It is clear that the details of Selection and Elimination Rules should depend upon both the information structure of the criterion pool, and the distribution of examinee ability.

Only three Selection Rules and three Elimination Rules were analyzed, and this analysis took place over only a single cycle of item pool refreshment. The rigid adherence to a fixed combination of a Selection rule with an Elimination rule over many cycles cannot be recommended. For example, if Elimination Rule 3, the elimination of the 25 most used items and the 5 least used items, were consistently used with any of the Selection Rules, over many

cycles of item pool refreshment the effective range of ability over which the adaptive test measures well would shrink and no appropriate replacement items would be available. In practice, it seems better to maintain flexibility, and to choose Selection Rules and Elimination Rules for the next refreshment of the item pool on an ad-hoc basis by frequent examination of item pool statistics as adaptive testing proceeds.

The Selection Rules studied all employ the screening of items on the basis of classical item statistics. This is more expensive than not screening items, as was done in the previous study, because of the necessary overproduction of items. The stricter the screening criteria, the greater the cost, as more of the items initially written will not meet the criteria. The benefit gained from the added expense of screening is the minimization of changes to the information structure of the item pool.

It seems clear that providing more items for seeding provides more flexibility in the choice of replacement items. This enhanced flexibility makes it easier to maintain the information structure of the pool, but it, too, incurs real-world costs. To collect the data for on-line calibration of more items requires, for a fixed number of examinees, that each examinee respond to more seeded items. If this is not feasible in terms of

lengthening examinee testing time, then more examinees are required. This lengthens the time required to collect the data for on-line calibration.

Eliminating over-exposed items from an adaptive test item pool seems a reasonable approach to maintaining test security. The definition of over-exposure used in this study was arbitrary -- the 25 most used items. No attempt was made to determine if this was reasonable. Other types of rules may function better in practice. For example, it may be more reasonable to set an absolute cut-off on the number of times an item can be administered before it is considered over-exposed.

The elimination of under-exposed items from the adaptive test item pool should be approached with caution, since this may reduce the effective range of the adaptive test. Careful consideration is needed to decide whether this reduction in range is tolerable, given the original purpose for which the adaptive test was constructed, and the potential benefits in terms of the information structure of the new pool.

For the particular context of this study, the results suggest two approaches to the choice of a selection rule combined with an elimination rule for a single cycle of item pool refreshment. One approach would eliminate only over-exposed items (Elimination

Rule 1) and choose replacements from 100 strictly screened seeded items (Selection Rule 3). A new item pool can be produced with almost the same information structure as the original pool. A second approach, one that should only be used with caution, would eliminate a small number of underexposed items also (Elimination Rule 3). Replacements chosen from 100 moderately screened seeded items (Selection Rule 2) can result in a new pool that is also very similar in information structure to the original pool.

This small study was not designed to examine a wide variety of selection and elimination rules in a variety of different contexts. However, based on the results, two more general conclusions are suggested:

1) Using conventional item statistics to screen items before deciding to seed them seems important and effective in terms of maintaining the information structure of the adaptive test item pool. The details of the screening criteria must depend upon the particular item pool and the examinees for whom the adaptive test is intended.

2) The on-line calibration of larger sets of seeded items from which to select replacements can substantially improve the ease with which the information structure of the pool can be

maintained by providing added flexibility in the choice of replacement items.

References

Bock, R. D., Davis, C. E., Holland, P. W., Levine, M. V., Samejima, F. & Stocking, M. L. (1988). Final report on on-line calibration. In preparation.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. Applied Psychological Measurement, 7, 201-210.

Wingersky, M. S. (1983). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In R. Hambleton (Ed.), ERIBC monograph on applications of item response theory (pp. 45-56). Vancouver, BC: Educational Research Institute of British Columbia.

Acknowledgment

This work was supported by Contract No. N00014-85-K-0683, project designation NR 150-539, from Cognitive Science Program, Cognitive and Neural Sciences Division, Office of Naval Research and Educational Testing Service. Reproduction in whole or in part is permitted for any purpose of the United States Government. The author wishes to acknowledge the invaluable assistance of Maxine B. Kingston in carrying out this study.

Table 1

Summary Statistics for Proportions-Correct and r-Biserials on the Item Sets Produced by the Three Selection Rules

Selection Rule 1, n = 50

	Mean	S.D.	Min	Max	Percentiles 10	25	50	75	90
proportion-correct	.58	.24	.21	.90	.27	.32	.63	.83	.87
r-biserial	.53	.16	.23	.73	.26	.40	.57	.66	.71

Correlation between p^+ and r-bis = .51

Selection Rule 2, n = 100

	Mean	S.D.	Min	Max	Percentiles 10	25	50	75	90
proportion-correct	.58	.23	.21	.90	.27	.34	.63	.80	.87
r-biserial	.54	.16	.23	.84	.30	.40	.58	.67	.72

Correlation between p^+ and r-bis = .55

Selection Rule 3, n = 100

	Mean	S.D.	Min	Max	Percentiles 10	25	50	75	90
proportion-correct	.61	.12	.42	.80	.44	.48	.63	.70	.78
r-biserial	.60	.13	.28	.84	.39	.52	.62	.68	.74

Correlation between p^+ and r-bis = .33

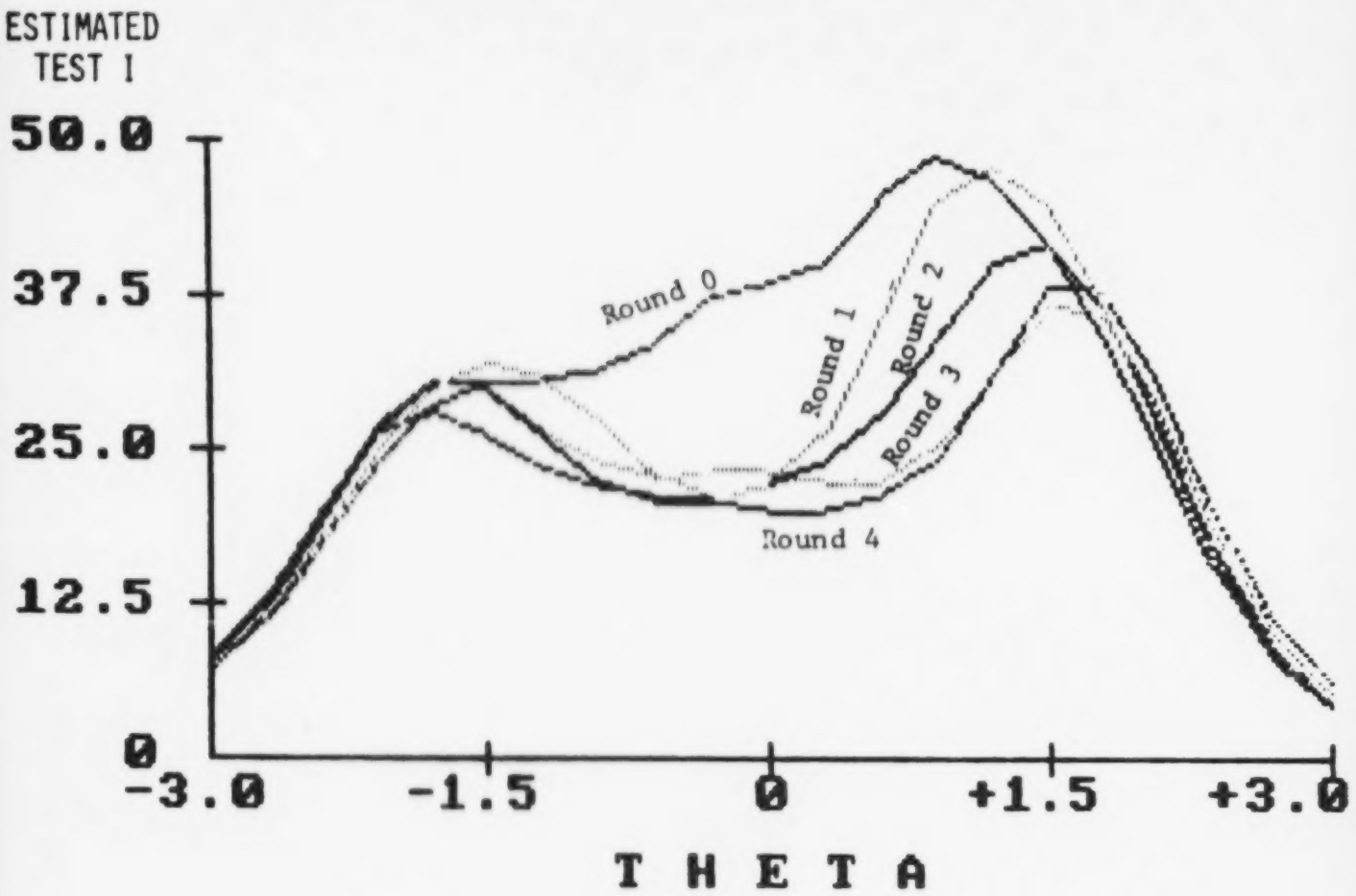

Figure 1. Estimated test information functions for the 100-item pools at each Round of the previous simulations for the LOGIST-based method of on-line calibration.

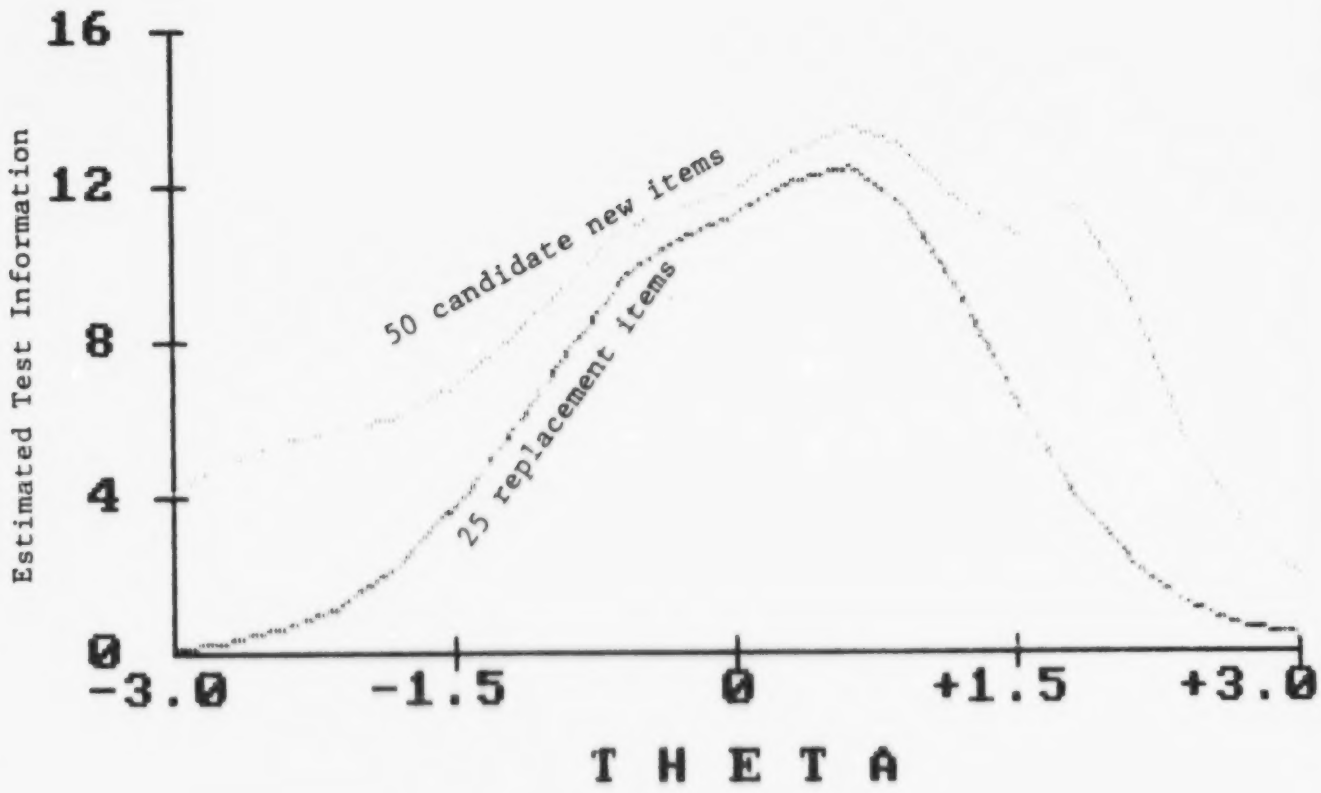

Figure 2. Estimated test information functions for the set of 50 randomly selected candidate new items and the 25 replacement items selected from the 50 for the refreshment of the Round 0 pool in the previous simulations.

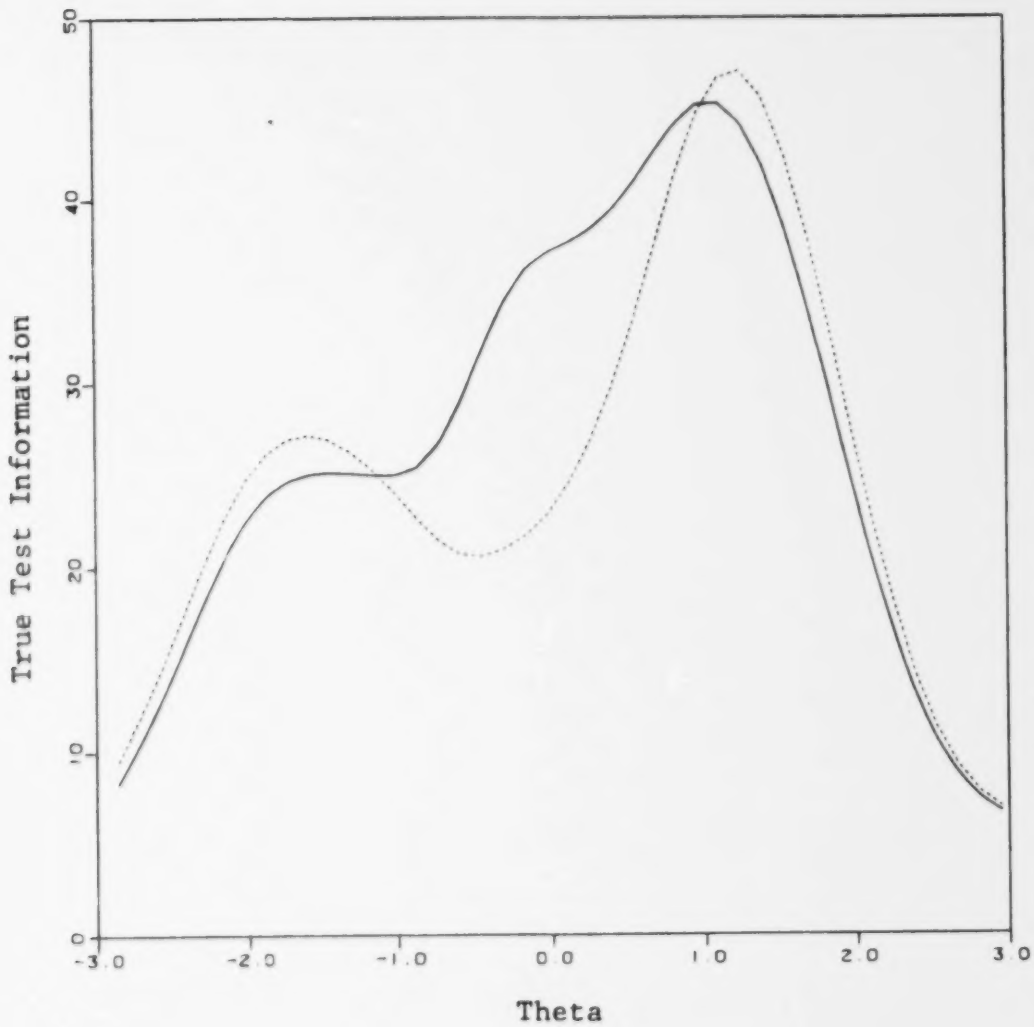

Figure 3. True test information functions for the Round 0 pool (solid line) and Round 1 pool (dotted line) from the previous simulations.

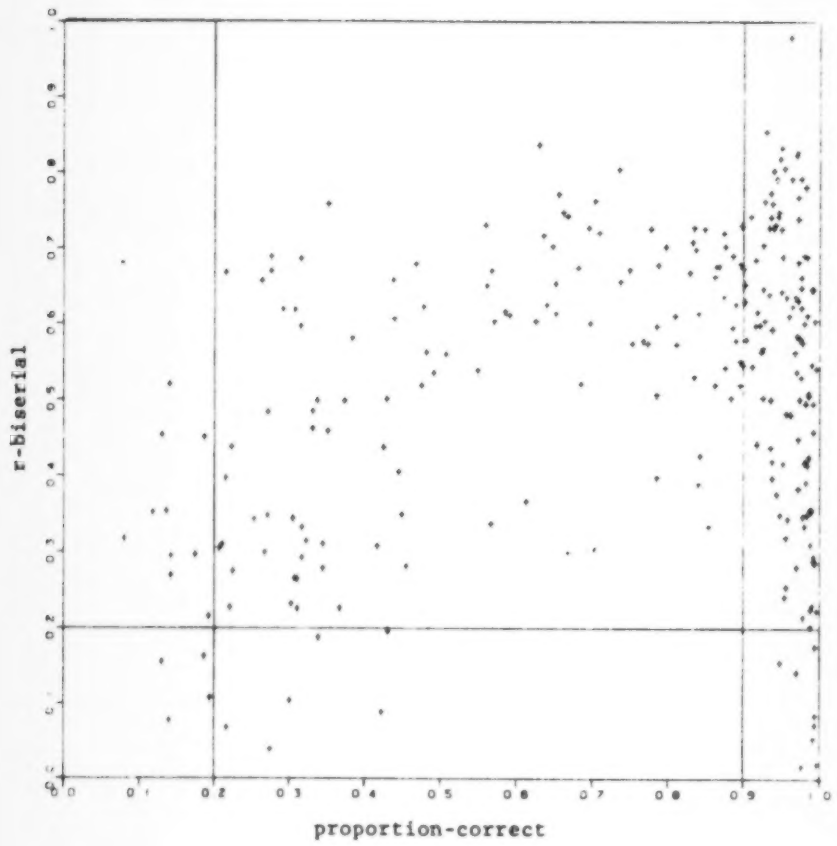

Figure 4. Scatterplot of r-biserials (vertical axis) and proportions-correct (horizontal axis) for the set of 258 items. The two vertical lines mark the less restrictive limits on proportions-correct for Selection Rules 1 and 2. The horizontal line marks the limit on r-biserials used for all three Selection Rules.

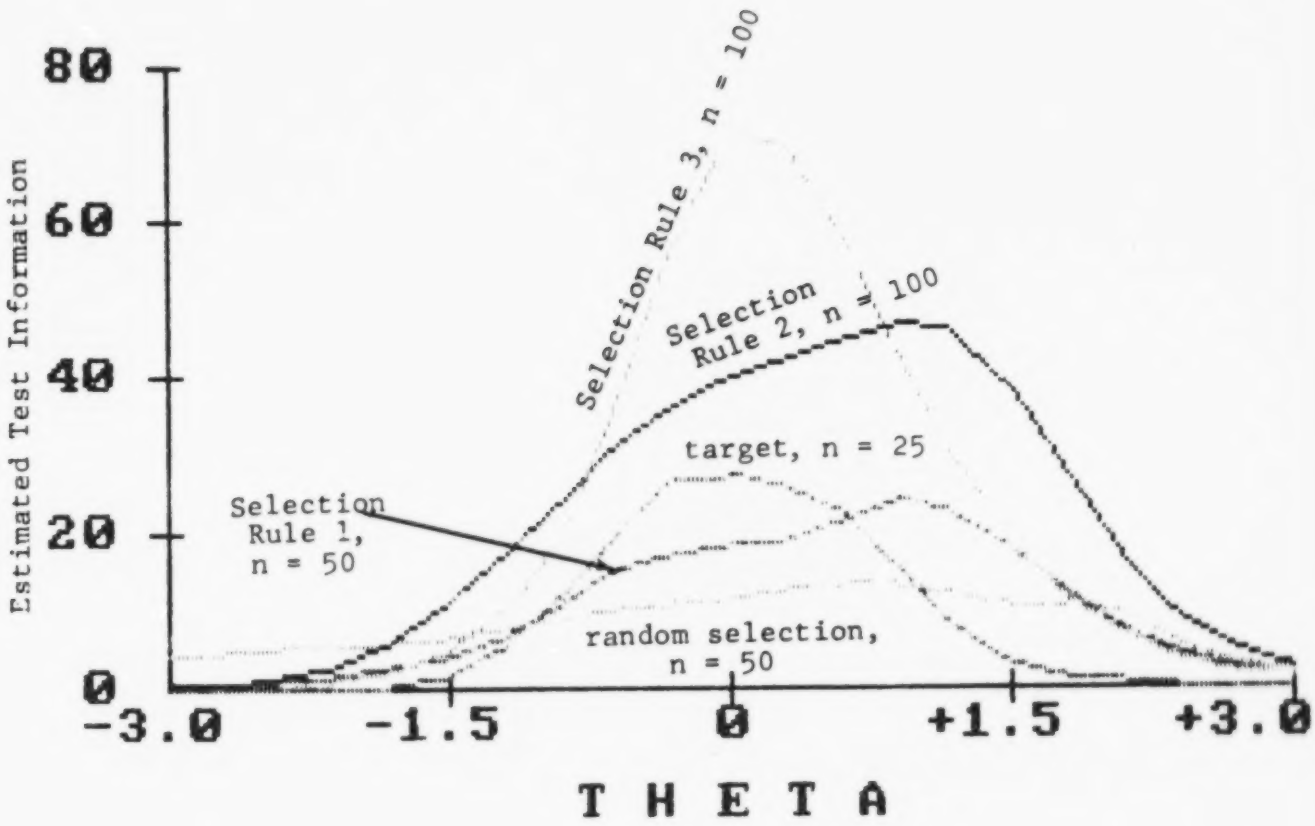

Figure 5. Estimated test information functions for the sets of seeded items resulting from the three Selection Rules of the current study and the random selection rule of the previous study. Also shown is the target test information function for the 25 most used items in the Round 0 item pool.

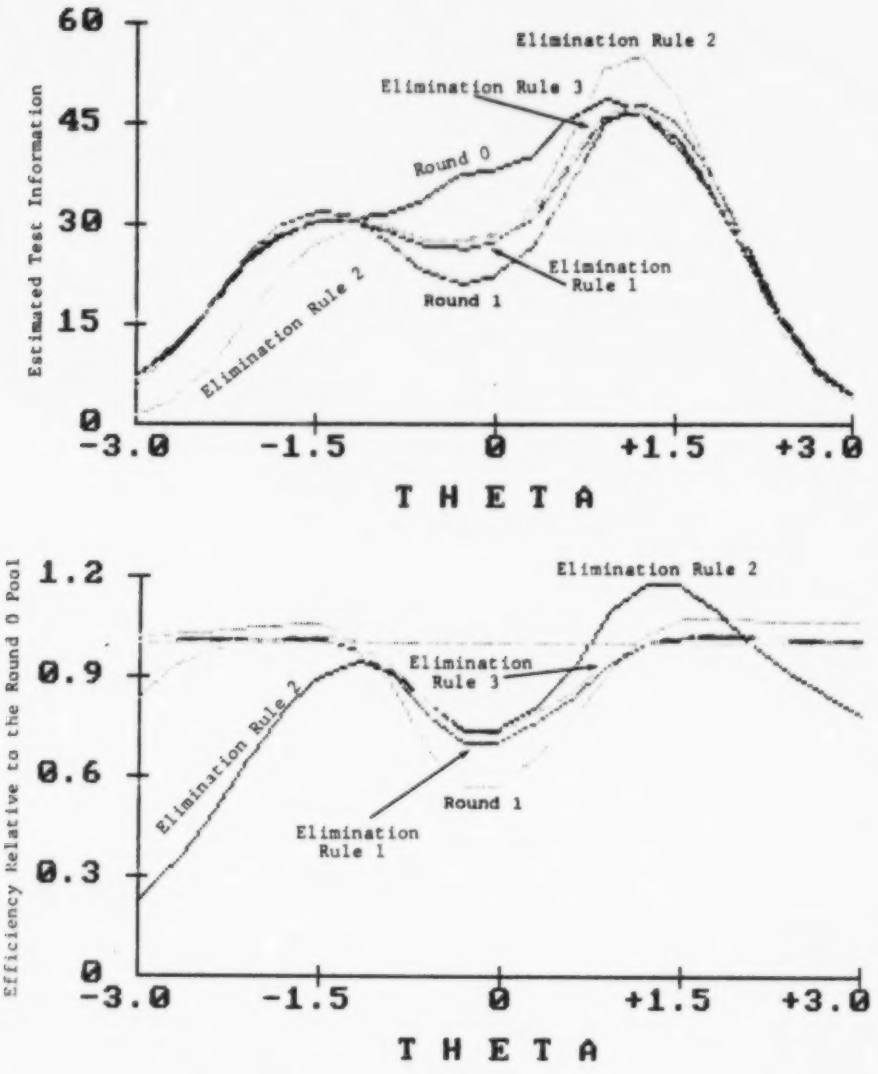

Figure 6. For Selection Rule 1: Estimated test information functions for the 100-item pools resulting from each Elimination Rule of the current study, the Round 1 pool of the previous study, and for the Round 0 pool (top); efficiency of each 100-item pool relative to the Round 0 pool (bottom).

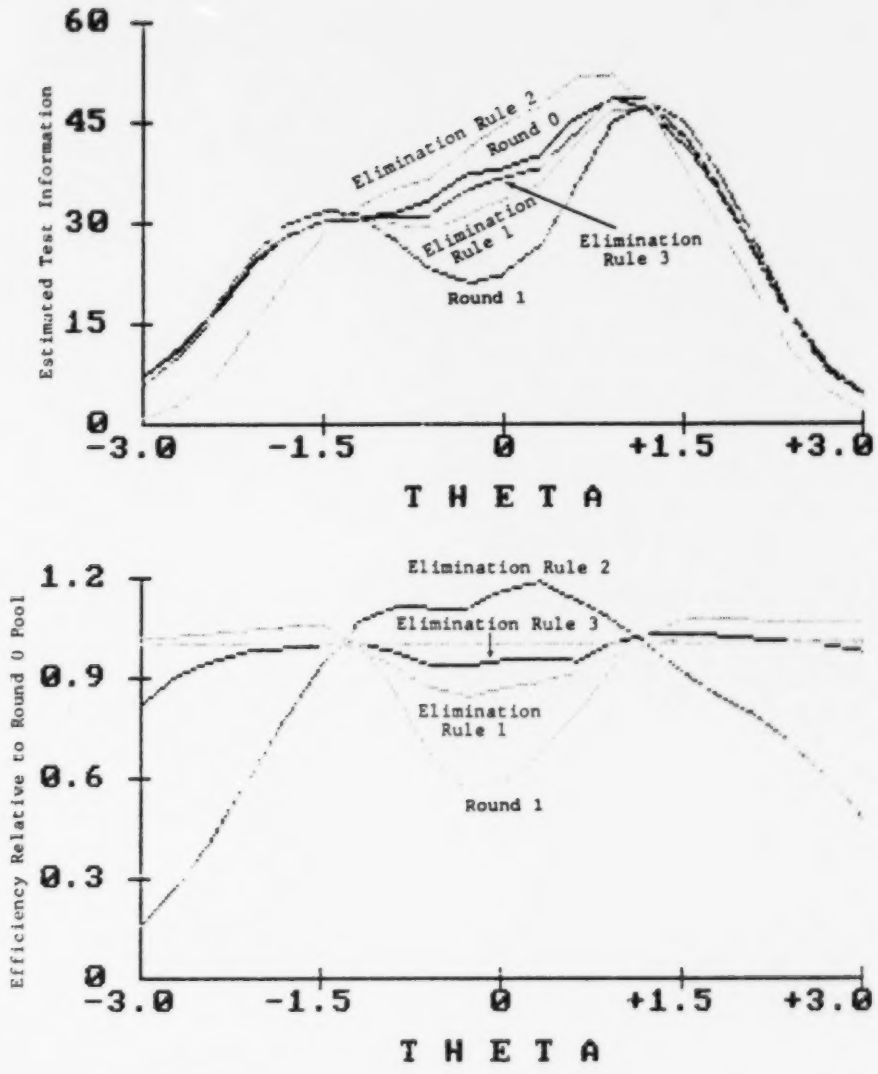

Figure 7. For Selection Rule 2: Estimated test information functions for the 100-item pools resulting from each Elimination Rule of the current study, the Round 1 pool of the previous study, and for the Round 0 pool (top); efficiency of each 100-item pool relative to the Round 0 pool (bottom).

BEST COPY AVAILABLE

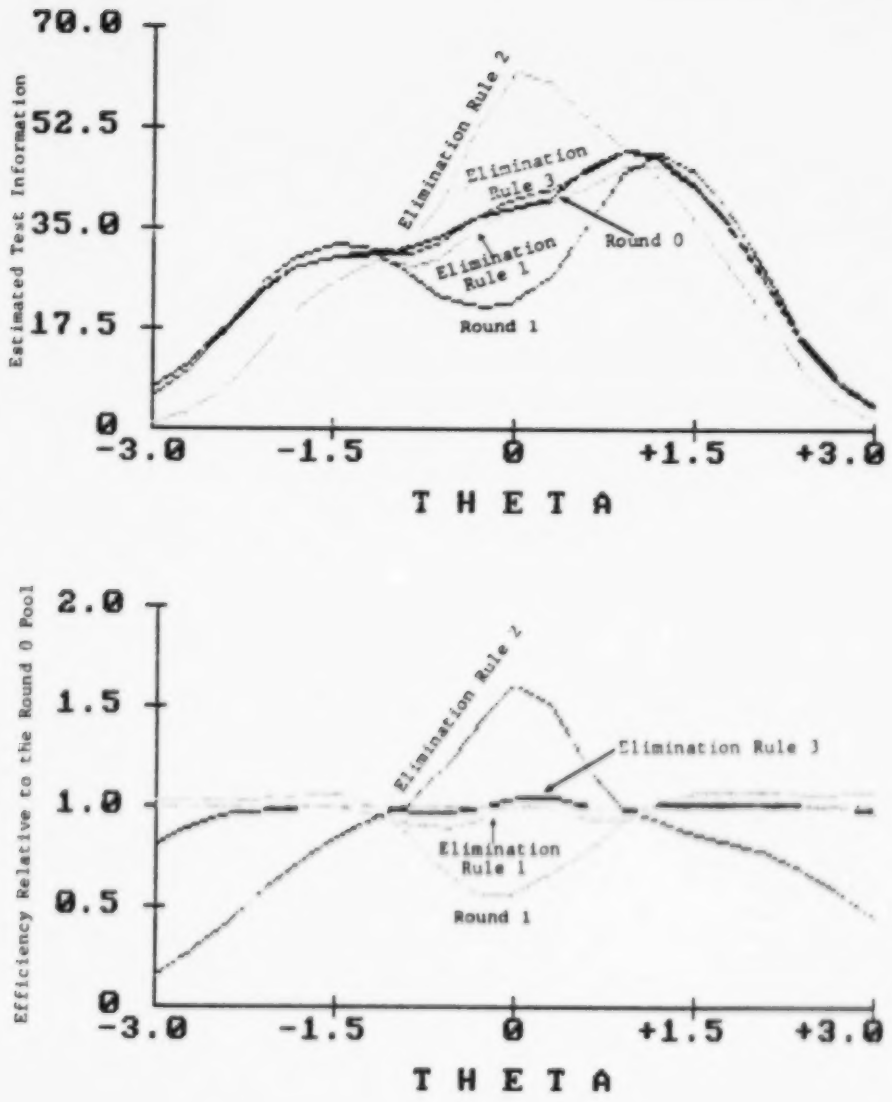

Figure 8. For Selection Rule 3: Estimated test information functions for the 100-item pools resulting from each Elimination Rule of the current study, the Round 1 pool of the previous study, and for the Round 0 pool (top); efficiency of each 100-item pool relative to the Round 0 pool (bottom).

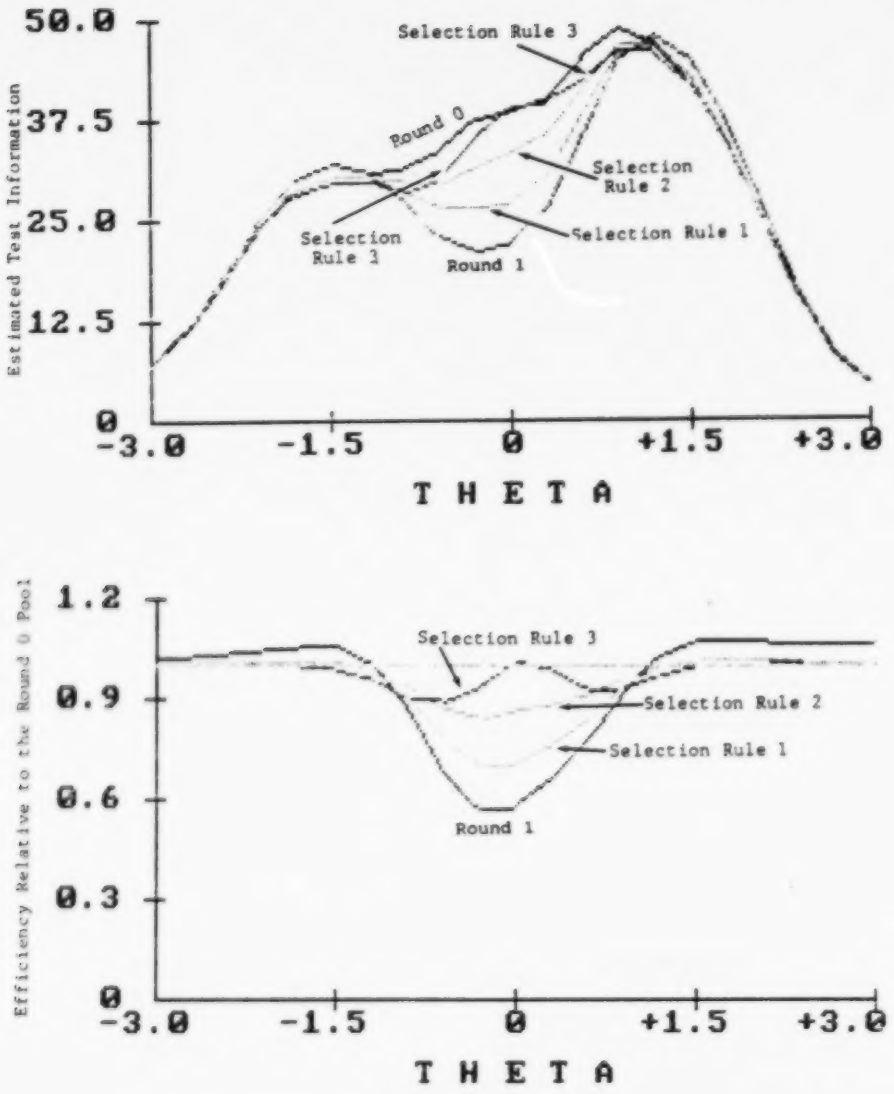

Figure 9. For Elimination Rule 1: Estimated test information functions for the 100-item pools resulting from each Selection Rule of the current study, the Round 1 pool rule of the previous study, and for the Round 0 pool (top); efficiency of each 100-item pool relative to the Round 0 pool (bottom).

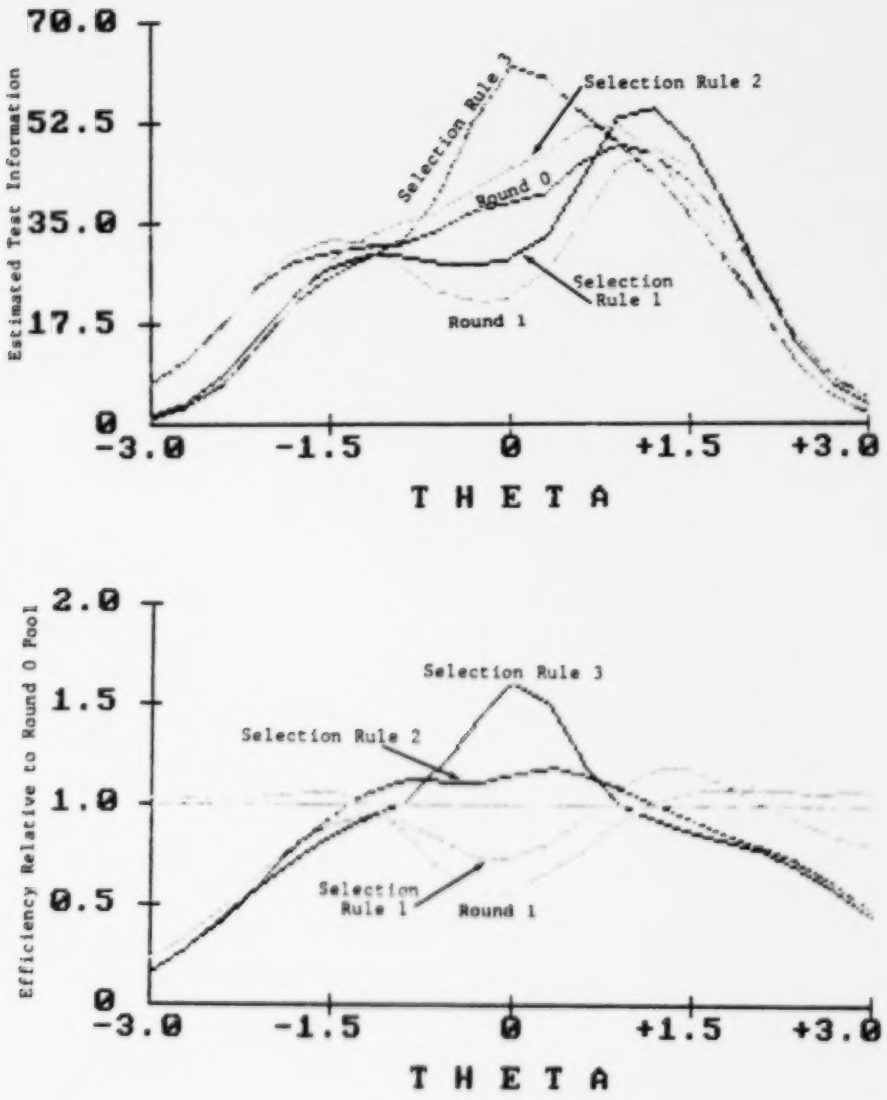

Figure 10. For Elimination Rule 2: Estimated test information functions for the 100-item pools resulting from each Selection Rule of the current study, the Round 1 pool of the previous study, and for the Round 0 pool (top); efficiency of each 100-item pool relative to the Round 0 pool (bottom).

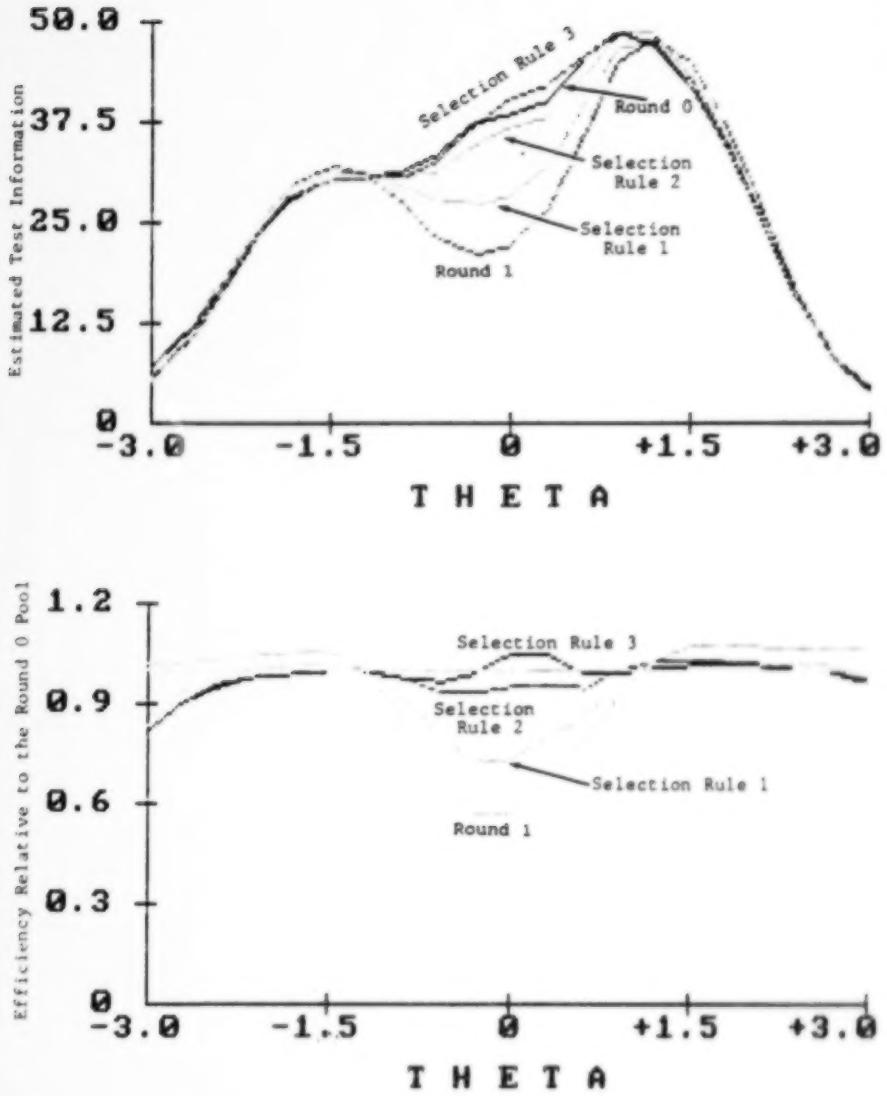

Figure 11. For Elimination Rule 3: Estimated test information functions for the 100-item pools resulting from each Selection Rule of the current study, the Round 1 pool of the previous study, and for the Round 0 pool (top); efficiency of each 100-item pool relative to the Round 0 pool (bottom).

Dr. Terry Ackerman
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. James Algina
1403 Norman Hall
University of Florida
Gainesville, FL 32605

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Mail Stop: 10-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N712
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Bruce Bloxom
Defense Manpower Data Center
550 Camino El Estero,
Suite 200
Monterey, CA 93943-3231

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code 7B
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd., North
Chapel Hill, NC 27514

Dr. Robert M. Carroll
Chief of Naval Operations
OP-01B2
Washington, DC 20350

Dr. Raymond E. Christal
UES LAMP Science Advisor
AFHRL/MOEL
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
Los Angeles, CA 90089-1061

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans F. Crombag
Faculty of Law
University of Limburg
P.O. Box 616
Maastricht
The NETHERLANDS 6200 MD

Dr. Timothy Davey
Educational Testing Service
Princeton, NJ 08541

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 4112
University of Maryland
College Park, MD 20742

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1K226
Piscataway, NJ 08854

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
2248 Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Englehard, Jr.
Division of Educational Studies
Emory University
210 Fishburne Bldg.
Atlanta, GA 30322

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. P-A. Federico
Code 51
NPRDC
San Diego, CA 92152-6800

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
P.O. Box 168
Iowa City, IA 52243

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
U.S. Army Headquarters
DAPE-MRR
The Pentagon
Washington, DC 20310-0300

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Alfred R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

DORNIER GMBH
P.O. Box 1420
D-7990 Friedrichshafen 1
WEST GERMANY

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hills South, Room 152
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Grant Henning
Senior Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 63
San Diego, CA 92152-6800

Dr. Paul W. Holland
Educational Testing Service, 21-T
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Dr. Paul Horst
677 G Street, #134
Chula Vista, CA 92010

Mr. Dick Hoshaw
OP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
3-104 Educ. N.
University of Alberta
Edmonton, Alberta
CANADA T6G 2G5

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Milton S. Katz
European Science Coordination
Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1500

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
Box 7246, Meas. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

Dr. Leonard Kroeker
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Dr. Robert L. Linn
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. George B. Macready
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. James R. McBride
The Psychological Corporation
1250 Sixth Avenue
San Diego, CA 92101

Dr. Clarence C. McCormick
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Dr. Robert McKinley
Educational Testing Service
16-T
Princeton, NJ 08541

Dr. James McMichael
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Barbara Means
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Headquarters Marine Corps
Code MPI-20
Washington, DC 20380

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Norman, OK 73071

Deputy Technical Director
NPRDC Code 01A
San Diego, CA 92152-6800

Director, Training Laboratory,
NPRDC (Code 05)
San Diego, CA 92152-6800

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

Director, Human Factors
& Organizational Systems Lab,
NPRDC (Code 07)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

Dr. James B. Olsen
WICAT Systems
1875 South State Street
Orem, UT 84058

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Office of Naval Research,
Code 125
800 N. Quincy Street
Arlington, VA 22217-5000

Assistant for MPT Research,
Development and Studies
OP 01B7
Washington, DC 20370

Dr. Judith Orasanu
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Randolph Park
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Department of Operations Research,
Naval Postgraduate School
Monterey, CA 93940

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MOA
Brooks AFB, TX 78235

Dr. Barry Riegelhaupt
HumRRO
1100 South Washington Street
Alexandria, VA 22314

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Dan Segall
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street, Suite 120
Alexandria, VA 22314-1713

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Richard C. Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Paul Speckman
University of Missouri
Department of Statistics
Columbia, MO 65201

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
Code-62
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
103 S. Mathews Avenue
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E. Street, NW
Washington, DC 20415

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 440
St. Paul, MN 55114

Dr. Frank L. Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Thomas A. Warm
Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Dr. Brian Waters
HumRRO
12908 Argyle Circle
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Ronald A. Weitzman
Box 146
Carmel, CA 93921

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 51
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90089-1061

German Military Representative
ATTN: Wolfgang Wildgrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
NRC MH-176
2101 Constitution Ave.
Washington, DC 20418

Dr. Martin F. Wiskoff
Defense Manpower Data Center
550 Camino El Estero
Suite 200
Monterey, CA 93943-3231

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
Code 51
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
03-T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550

Mr. Anthony R. Zara
National Council of State
Boards of Nursing, Inc.
625 North Michigan Avenue
Suite 1544
Chicago, IL 60611

Dr. Peter Stoloff
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268